// ADITYA SHARMA

(778)-917-8028|asa498@sfu.ca|LinkedIn

CORE SKILLS:

Databricks & Cloud Data Engineering: Databricks Lakehouse, Delta Lake, Databricks SQL, MLflow, Apache Spark, PySpark, Azure Data Factory, Data Migration, Structured Streaming, Data Governance
ETL & Data Pipeline Optimization: Apache Airflow, SQL, SparkSQL, API Integration, SFTP Data Ingestion, Feature Engineering.
Data Warehousing & Performance Tuning: Data Modeling, Query Optimization, Partitioning, Indexing.

CERTIFICATIONS:

Databricks Data Engineer Associate Certification (Issued: March, 2025)

PROFESSIONAL EXPERIENCE:

BIG DATA SUPPORT SPECIALIST Sep 2023 – Apr 2024
Simon Fraser University|Vancouver

- **Designed, implemented, and maintained scalable ETL pipelines** using **Python, Apache Spark, and Google BigQuery,** processing millions of records with **99.9% uptime.**
- **Automated data** ingestion from SFTP servers and REST APIs, enhancing **real-time data processing capabilities** by **30%.**
- **Developed and optimized SQL-based schemas** for internal data warehouses, improving query efficiency by **40%.**
- **Monitored and troubleshot** operational issues in ETL workflows, ensuring data consistency and minimal downtime.

DATA ENGINEER INTERN Sep 2020 – Jul 2021
Indian Council of Medical Research|New Delhi

- Built and optimized ETL pipelines with **Python and SQL**, **processing medical data** for large-scale epidemiological studies.
- Designed data models to improve vaccine efficacy analysis **by 30%**, ensuring compliance with **data governance policies.**
- Resolved **data pipeline failures** by analyzing root causes, reducing **incident response time by 25%.**
- Collaborated with stakeholders to create dashboards and reports, driving actionable insights for public health strategies.

JUNIOR DATA ANALYST Nov 2019 – Dec 2019
NuWave Solutions pvt ltd|New Delhi

- Optimized customer segmentation by analyzing **5 million** transaction records using **K-Means clustering** in Python, leading to a **20%** increase in targeted marketing effectiveness and customer retention.
- Enhanced demand forecasting with an **ensemble regression model** in Python, **achieving 92% sales prediction accuracy** while ensuring **data confidentiality and integrity during processing.**

PROJECTS:

Analytical Study of Mental Well-being Among University Students Jan 2024 – Apr 2024

- Conducted a **secure analysis** of university students' mental well-being **using OLS regression** and **Mixed Linear Models**, identifying factors that improved mental health outcomes by **15%,** particularly in reducing stress levels.
- Processed and cleaned data using **Python and Pandas**, then created **dynamic Tableau visualizations** that revealed a positive correlation between regular exercise and reduced anxiety, boosting stakeholder comprehension by **20%** and driving actionable recommendations for mental health initiatives.

Data Driven Customer Segmentation and game recommendations for SFU Jan 2023 – Apr 2023

- Developed an end-to-end **ETL pipeline** for web scraping, cleaning, and storing gaming data in **BigQuery**, which resulted in 89% recommendation accuracy and 40% improved student engagement.
- Performed feature engineering using **SQL** and **dimensionality reduction with PCA**, achieving **89%** accuracy in game recommendations and enhancing predictive model performance.
- Developed a **decision tree classifier** using **scikit-learn** to forecast game success with an **84%** F1-score, and created interactive visualizations using Tableau, which increased stakeholder engagement by **40%** and revealed the impact of game genre and developer on success.

EDUCATION:

Master's of Computer Science in Big Data Sep 2022 – Apr 2024
Simon Fraser University **Burnaby B.C**
Bachelor's in Technology in Computer Science Jul 2018 – May 2022
SRM Institute of Science and Technology **Chennai India**

SUMMARY:

Data Engineer specializing in Databricks Lakehouse, Delta Lake, and Apache Spark for scalable ETL pipelines and big data optimization. Proficient in Databricks SQL, structured streaming, and Airflow for batch and real-time processing. Skilled in cloud automation, data modeling, and query tuning to enhance efficiency in high-volume environments.